

Data Science Hackathon

Equipe SmallData





Equipe SmallData

Bruno Henrique Meyer

Daniel Dutra Rufasto

Francine Machado Resende

Gabriel Olescki

INFORMÁTICA BIOMÉDICA - UFPR



Abordagem escolhida

Como utilizar a informação textual para prever o índice de satisfação?



Motivação

Textos de avaliações livres em outros sites que não necessariamente possuem nota

Análise da reputação da marca ou do modelo.

Ex: Facebook e Twitter



Metodologia

Linguagem trabalhada: Python

Script para normalização dos arquivos de dados (parse)

Extração das informações relevantes para o problema

- Avaliações escritas e índice de satisfação (nota de recomendação)



Metodologia

Processamento de Linguagem Natural

- Agrupamento de palavras relevantes
- Seleção de palavras discriminantes (variância da frequência)
- Montagem do vetor de características



Resultados

Faltou tempo :(